Polarization Measurement through Ordered Latent Class Analysis

Boris Sokolov
LCSR HSE, Junior Research Fellow,
SPbU, Postgraduate Student

bssokolov@gmail.com

LCSR International Workshop,
April 3, 2014
Moscow

Why polarization?

- Polarization refers to level of diversity in society on some specific dimension.
- Polarization also reflects a conflict potential caused by diversity.
- Attitudinal polarization is an evidence of cultural cleavage (e.g. so called ‘modernization’ cleavage assumed by ‘losers of modernization’ thesis)
- Attitudinal polarization may be used as a second-level predictor for analyses of many social processes, especially related to politics and ethnic relations.
- Polarization (and related cleavages) may be interesting to model as well.

Measurement of Polarization Previous developments

- Variance (or Standard Deviation)
- Kurtosis
- Foster-Wolfson Index
- Duclos-Esteban-Ray family of indices
- Ethno-Linguistic Fractionalization Index
- Reynal-Querol Index of polarization
- Various measures of ordinal variation
- Visual distribution comparisons
- *Ad hoc* methods (like Mouw and Sobel 2001)

Polarization in Survey Data

- The main objects of interest are latent constructs (measured through multiple manifest variables).
- Information about distributional parameters of latent variables provided by relevant statistical software is limited.
- Measuring polarization for aggregated factor scores seems to be an inaccurate approach due to possible non-normality, multidimensionality, and measurement non-equivalence of latent scale.

Why (Ordinal) Latent Classes?

- LCA may easily **handle non-normality** of latent variable
- LCA **allows for multidimensionality**: when the latent categorical variable is nominal rather than ordinal, it is impossible to order all individuals on all items in the same direction.
- LCA allows for testing **measurement Invariance**
- LCA **provides unique observed indicator for latent variable** by classifying respondents according to their value patterns. Several existing ordinal measures of polarization are easily applicable to the resulting classification

Latent Class Model

- X1, X2, X3, and X4 are observed variables
- Y a latent categorical variable which accounts for the relationships among these four observed variables
- $\pi_{ijklt}^{X1X2X3X4Y} = \pi_t^Y \pi_{it}^{X1|Y} \pi_{jt}^{X2|Y} \pi_{kt}^{X3|Y} \pi_{lt}^{X4|Y}$
- π_t^Y is a probability that a randomly selected individual will be in latent class t of latent variable Y
- $\pi_{it}^{X1|Y}$ is a probability that a member of latent class t will choose a response category i for observed item $X1$
- $\pi_{jt}^{X2|Y}$ is a probability that a member of latent class t will choose a response category j for observed item $X2$
- $\pi_{kt}^{X3|Y}$ is a probability that a member of latent class t will choose a response category k for observed item $X3$
- $\pi_{lt}^{X4|Y}$ is a probability that a member of latent class t will choose a response category l for observed item $X4$

Ordinal Latent Classes

- Ordering of the categories of the latent variable is provided by imposing inequality constraints on model parameters: means for continuous manifest variables and thresholds for binary and ordinal manifest variables.
- In MPLUS, thresholds τ_{it} are used instead probabilities $\pi_{it}^{Xn|Y}$ (logistic parameterization of LCA model)
- Large positive thresholds indicate that (cumulative) probability of a specific response value is relatively low, whereas large negative values suggest that the probability of the response is relatively high.

- Inequality constraint $\boldsymbol{\tau_{i1} < \tau_{i2} < \tau_{i3} < \tau_{i4}}$ assumes the following ordering of classes for threshold τ for variable i : Class 1 > Class 2 > Class 3 > Class4

Approach to the Measurement of Polarization

Step 1. Selecting a model with an optimal number of latent classes. Best model must satisfy three following requirements

- 1) be parsimonious: model with K classes should not include classes which are subgroups of classes identified in a model with K - 1 latent categories.
- 2) be almost ordinal: include very few parameters violating class-ordering
- 3) show the best fit (aBIC and BLRT) comparing to all other models which satisfy 1) and 2)

Step 2. Testing for ordinality (*unidimensionality, or strict monotonicity*) of latent trait: comparing unconstrained and strictly ordered models. Order-constrained hypothesis is tested directly by using Bayes factor approach

Step 3. Applying relevant index of nominal or ordinal polarization (depending on the results from the Step 2) to class proportions for each country obtained at the first stage.

Bonus. Exploring measurement invariance and cross-country differences in class proportions

Polarization Indices

- Reynal-Querol Index (nominal)
- Standardized Van der Eijk's Agreement A measure
- Berry/Mielke Index of Ordinal Variation
- Leik's Ordinal Variation Index
- L-Squared

Data

- Survival/Self-Expression Values. WVS, Fifth Wave
- Manifest variables 1: Happiness, Tolerance for Homosexuality, Trust, Four-Item Postmaterialism Index (as a single variable), Signing Petition
- Manifest variables 2: Tolerance for Homosexuality. Four-Item Postmaterialism, Signing Petition
- 29 European Countries: 27 EU members, Norway, and Switzerland
- 42817 respondents
- Data were not weighted
- Data were not imputed

Fit Statistics for Competing Models

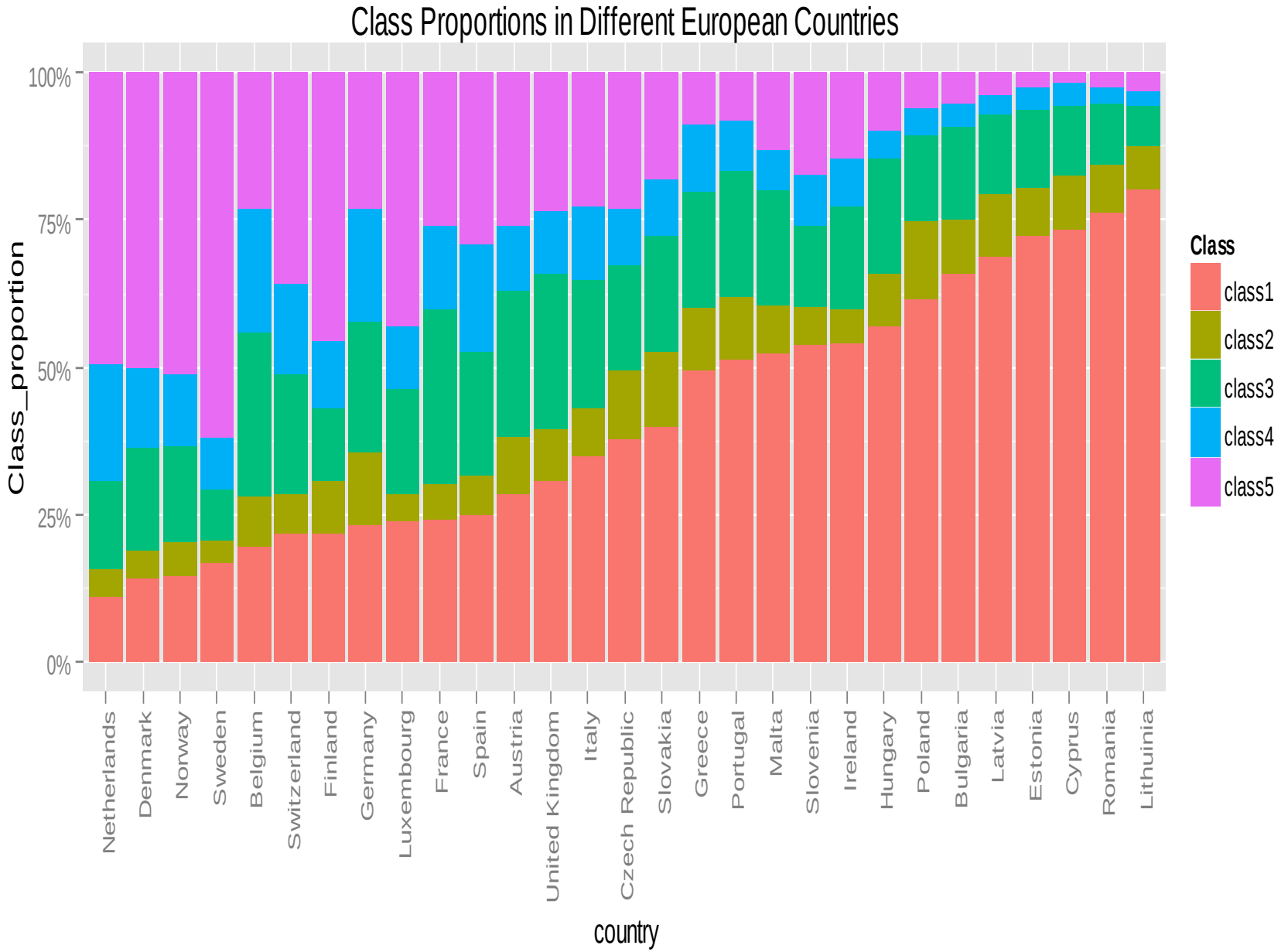
	aBIC	LMR Test p-value	BLRT p-value	Free Parameters	Violations of Ordering
Three Classes	471413.682	0.000	0.000	30	0
Four Classes	463097.672	0.000	0.000	40	1
Five Classes	448977.323	0.000	0.000	50	1
Five Classes_Ord	448977.441	***	***	50	0
Six Classes	446572.609	0.000	0.000	60	3
Six Classes_Ord	469077.511	***	***	55	0
Seven Classes	444052.829	0.000	0.000	70	6
Seven_Classes_Ord	***	***	***	64	0

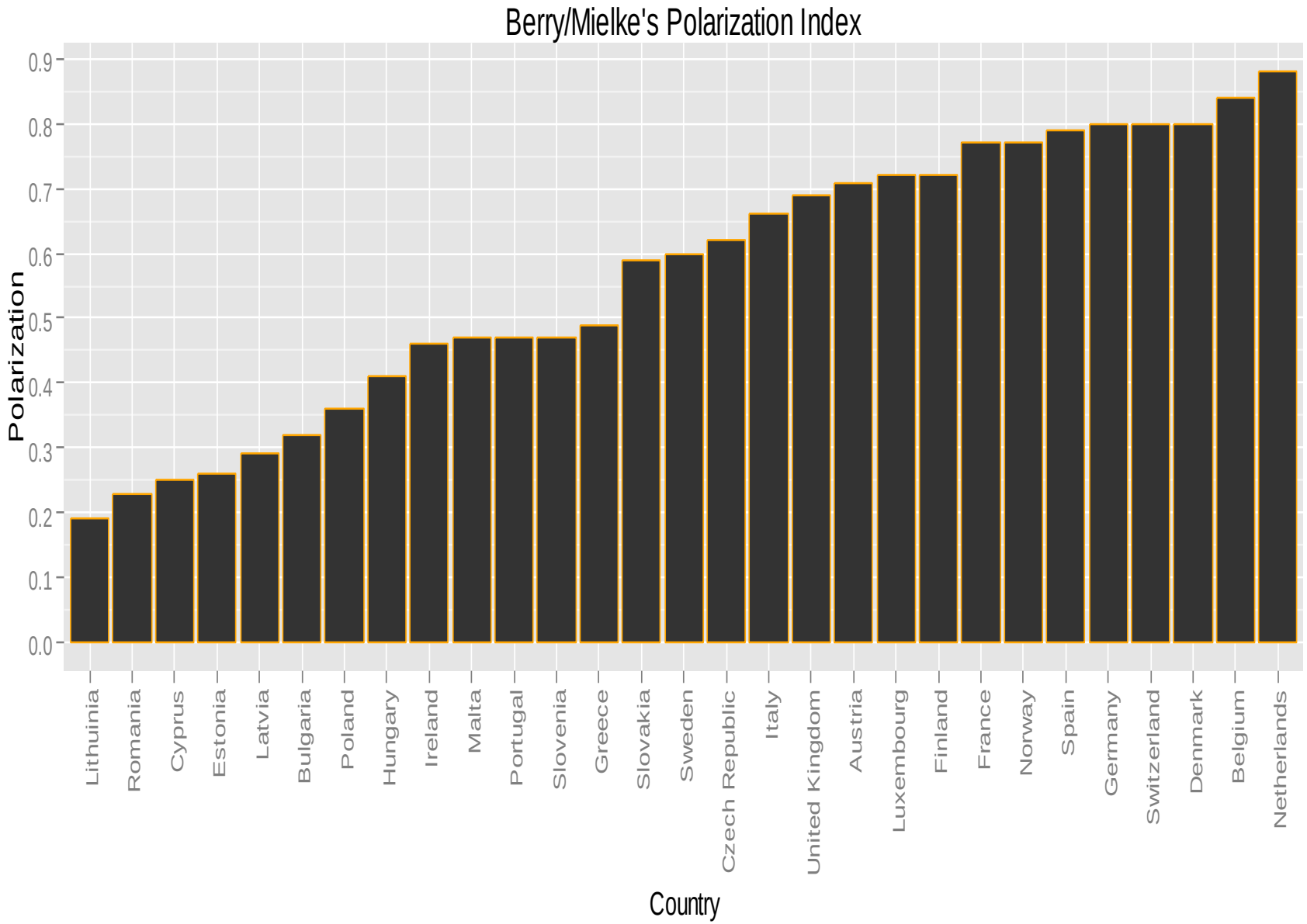
Thresholds and Means Estimates for the Five-Class Unconstrained Model

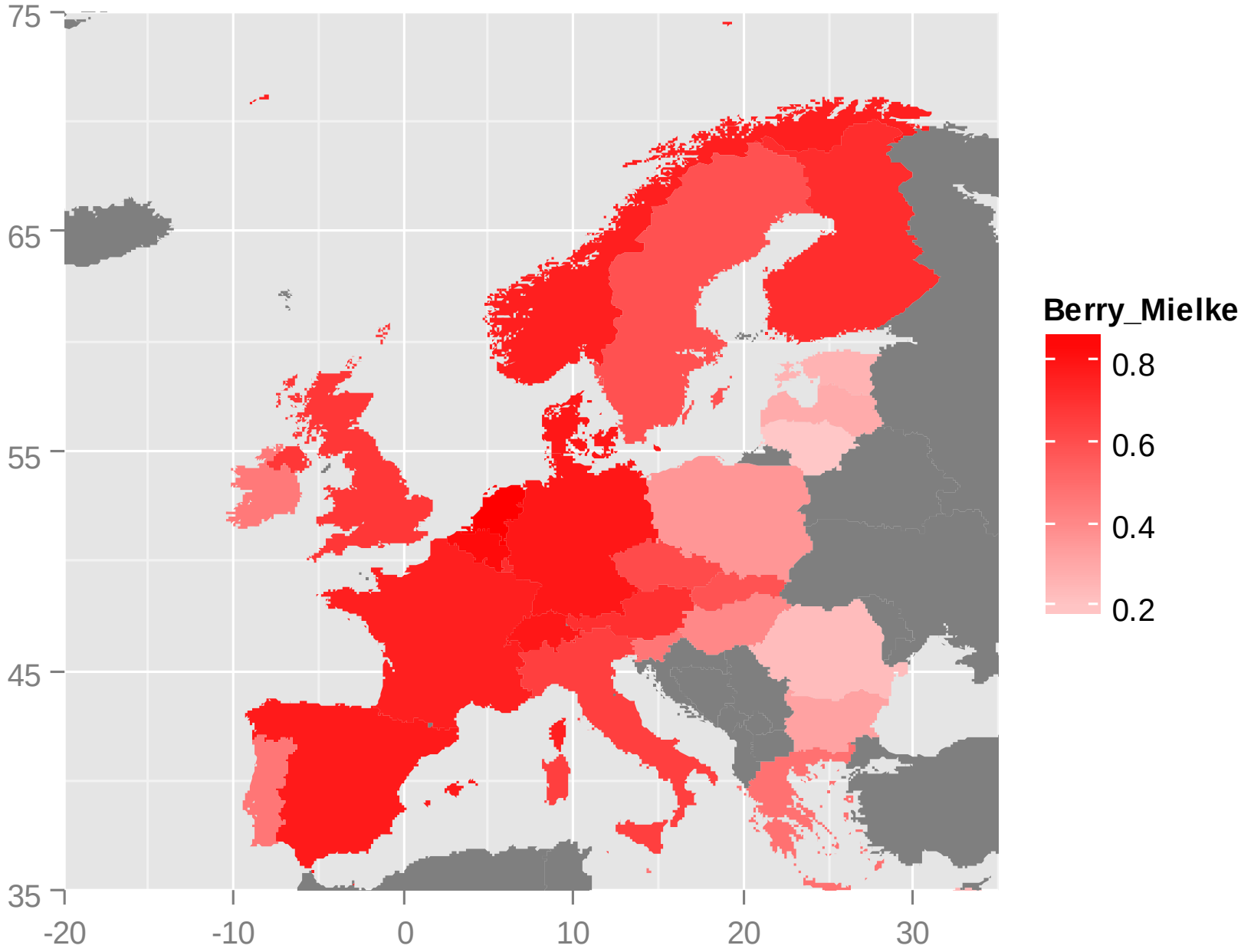
	Happy1	Happy2	Happy3	Pmat1	Pmat2	Trust	Petition 1	Petition 2	Homosex	Order
Class1	-3.84	-1.692	1.348	-0.836	2.365	0.859	-0.798	0.735	3.452	2
Class2	-3.495	-1.278	1.357	-0.598	2.631	1.174	-0.231	1.068	1.118	1
Class3	***-4.492***	-2.433	0.604	-1.706	1.291	-0.208	-2.322	-0.723	9.819	5
Class4	-4.3	-1.931	0.998	-1.038	1.991	0.69	-1.189	0.263	5.256	3
Class5	***-4.555***	-2.084	0.865	-1.294	1.661	0.394	-1.637	-0.134	7.557	4

Thresholds and Means Estimates for the Five-Class Model with Inequality Constraints

	Happy1	Happy2	Happy3	Pmat1	Pmat2	Trust	Petition 1	Petition 2	Homosex	Order
Class1	-3.84	-1.692	1.348	-0.836	2.365	0.859	-0.798	0.735	3.452	2
Class2	-3.495	-1.278	1.357	-0.598	2.631	1.174	-0.231	1.068	1.118	1
Class3	***-4.513***	-2.433	0.604	-1.706	1.291	-0.208	-2.322	-0.723	9.819	5
Class4	-4.3	-1.931	0.998	-1.038	1.991	0.69	-1.189	0.263	5.256	3
Class5	***-4.512***	-2.084	0.865	-1.294	1.661	0.394	-1.637	-0.134	7.557	4


Class Proportions in Different European Countries
Class_proportion
100%
75%
50%
25%
0%
Netherlands
Denmark
Norway
Sweden
Belgium
Switzerland
Finland
Germany
Luxembourg
France
Spain
Austria
United Kingdom
Italy
Czech Republic
Slovakia
Greece
Portugal
Malta
Slovenia
Ireland
Hungary
Poland
Bulgaria
Latvia
Estonia
Cyprus
Romania
Lithuinia
country
Class
class1
class2
class3
class4
class5


Berry/Mielke's Polarization Index
Polarization
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Lithuinia
Romania
Cyprus
Estonia
Latvia
Bulgaria
Poland
Hungary
Ireland
Malta
Portugal
Slovenia
Greece
Slovakia
Sweden
Czech Republic
Italy
United Kingdom
Austria
Luxembourg
Finland
France
Norway
Spain
Germany
Switzerland
Denmark
Belgium
Netherlands
Country


75
65
55
45
35
-20
-10
0
10
20
30
Berry_Mielke
0.8
0.6
0.4
0.2

Polarization Patterns for Five-Class Five-Item Model

- Class proportions vary in a large amount between countries
- There is a clear pattern: Eastern European countries shows larger proportions of survival classes (that is, less “modernized” classes)
- The less polarized countries are at the same time the less modernized while many developed countries are highly polarized
- Modernization and spread of self-expression values lead to the growth of value polarization?

Investigating the latent trait underlying the survival/self-expression values

- For five-item models, strict unidimensionality (class ordering) holds only for models with no more than five classes. For three-item models even nine-class solution is plausible.
- When the number of classes is relatively large (to approximate continuous distribution), the distribution of latent trait is trimodal, which indicates non-normality of the self-expression index.
- Country-by-country analysis shows that the class ordering identified in five-class five-tem solution is not robust across countries. Therefore, it is likely that configural measurement invariance does not hold for categorical representation of self-expression values index.
- Surprisingly, class ordering is more frequently violated in Western European countries, rather than in less developed post-communist or southern European societies.

Shortcomings and limitations

- Trade-off between efficiency and computational time might lead to biased parameter estimates
- Measurement invariance was not tested in a formal way
- LCA model selection may seem quite arbitrary

Further development

- Bayesian LCA
- Testing for local homogeneity in IRT framework instead of LCA measurement invariance
- Adding covariates
- Any advice is highly welcomed!!

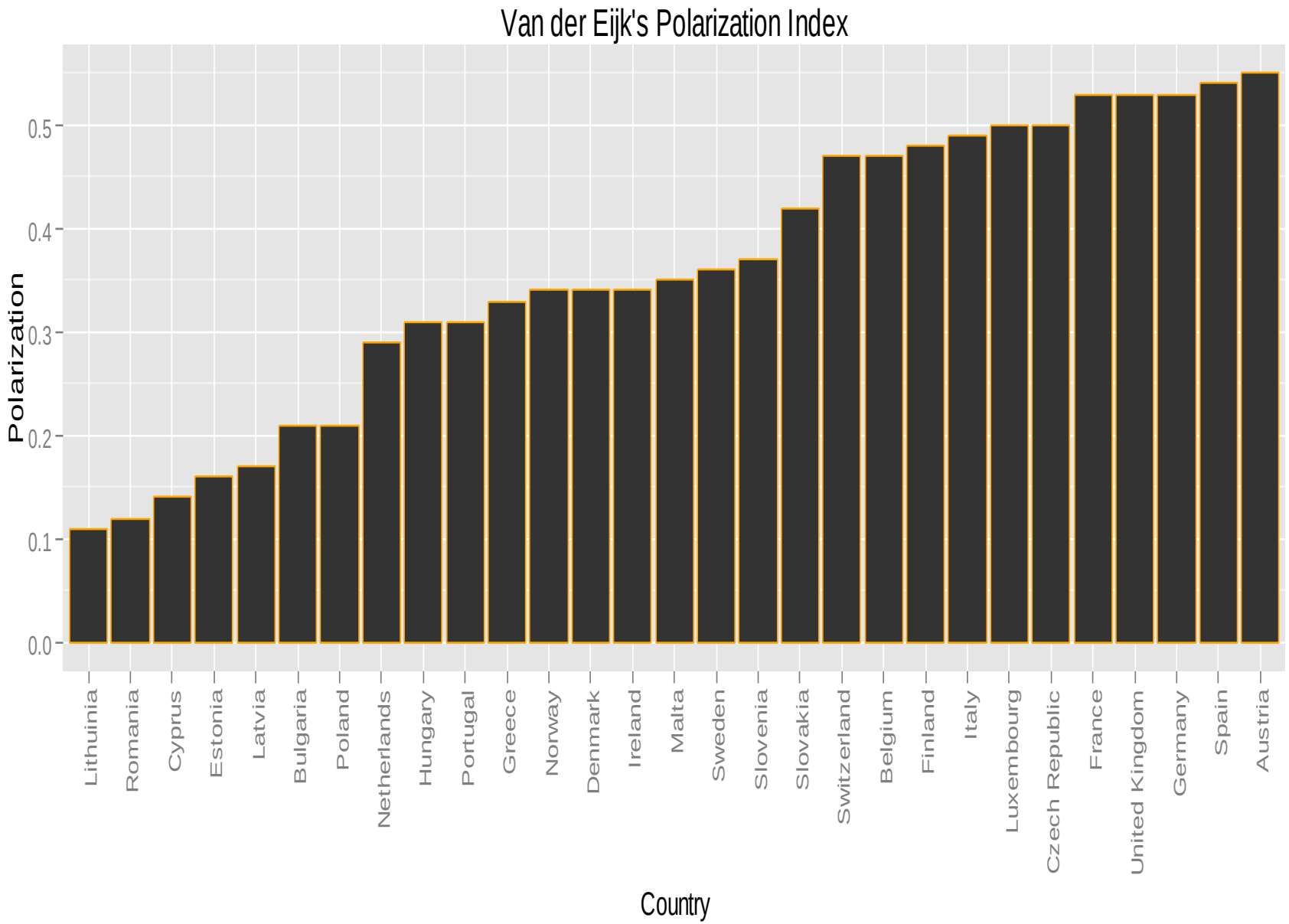
Thank you very much
for your attention!

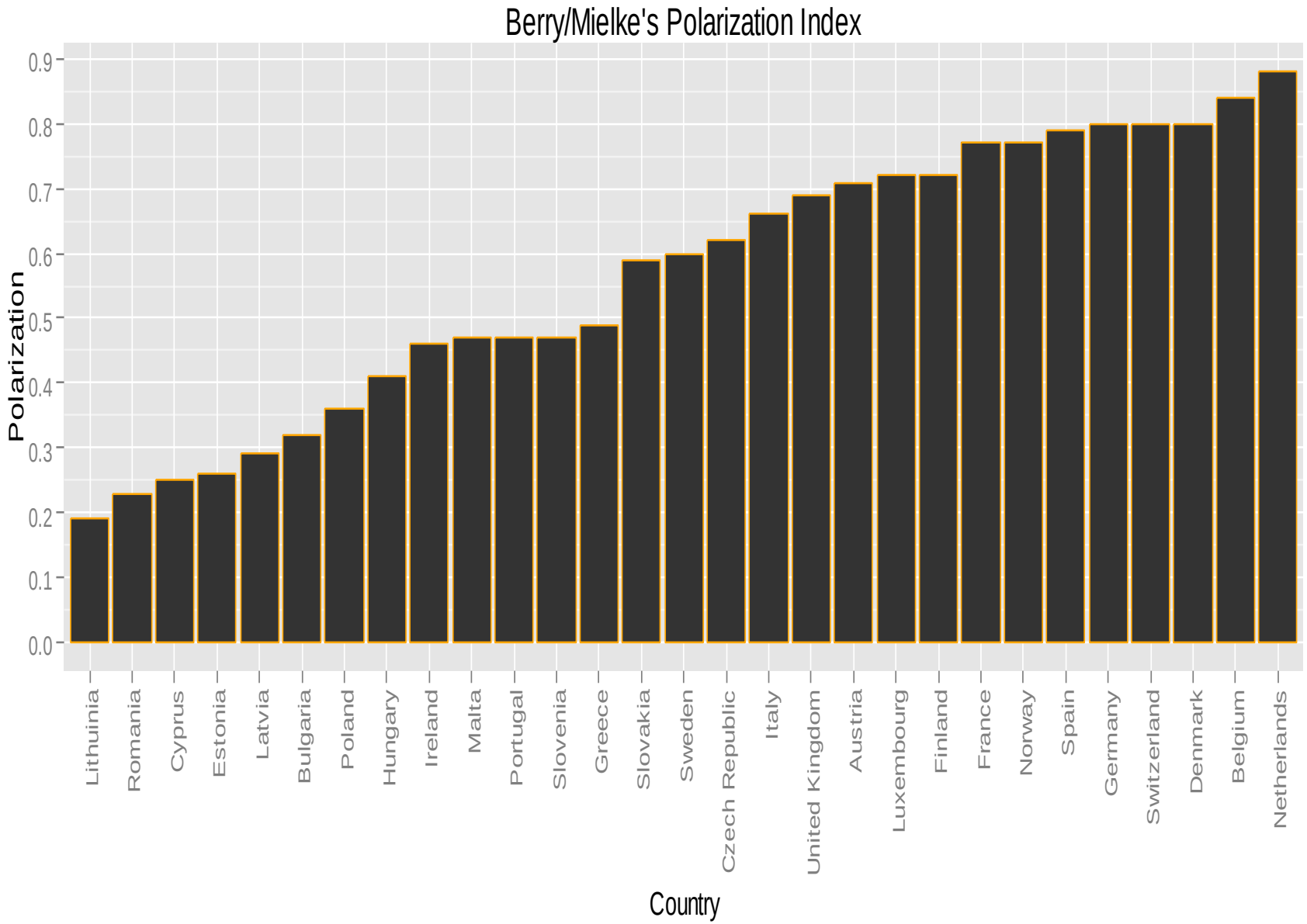
Thresholds and Means Estimates for the Six-Class Unconstrained Model

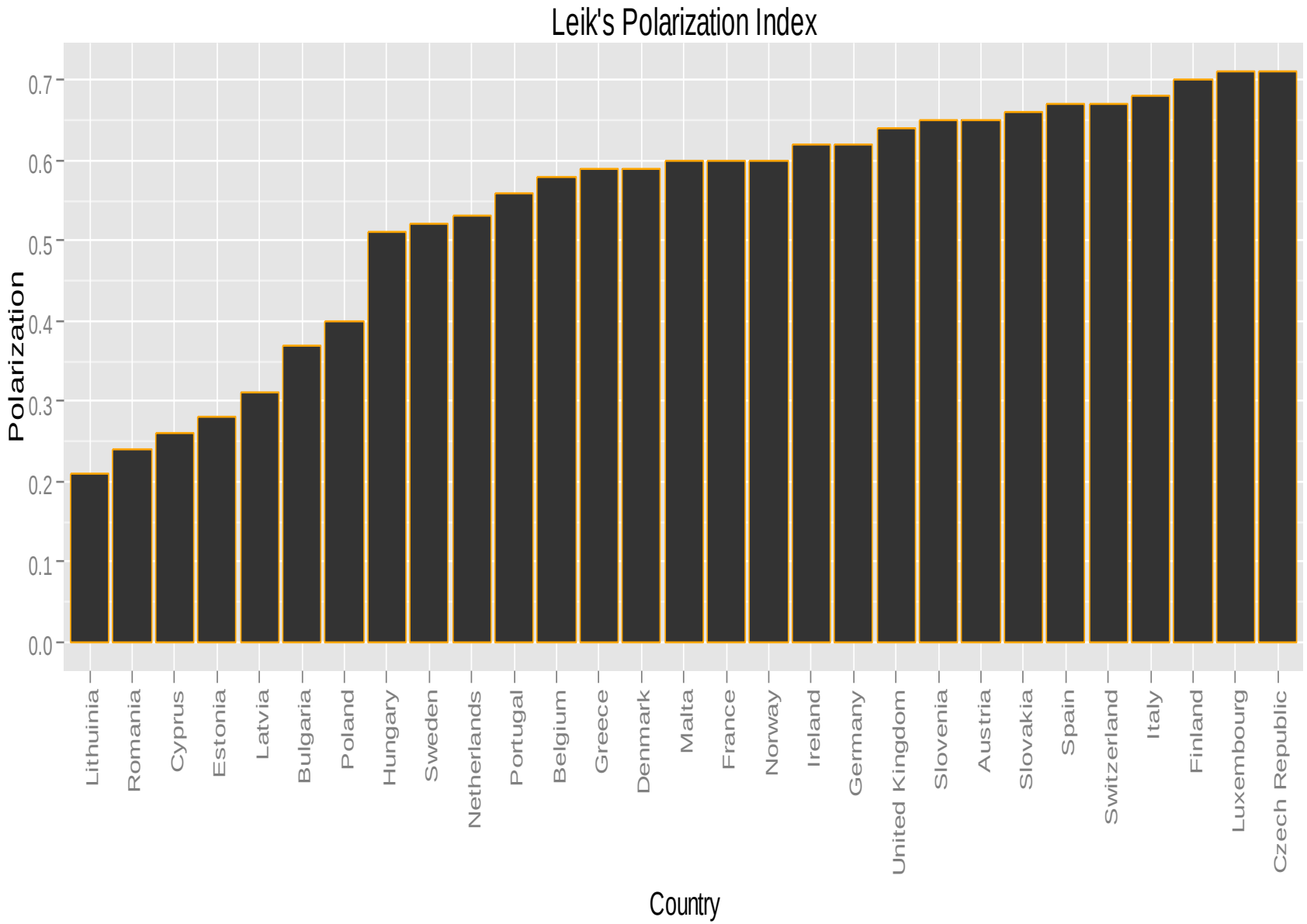
	Happy1	Happy2	Happy3	pmat1	pmat2	trust	petition 1	petition 2	homosex	Order
Class1	-3.83	-1.682	1.375	-0.825	2.393	0.865	-0.778	0.763	3.426	2
Class2	***-4.492***	-2.43	0.605	-1.705	1.292	-0.207	-2.321	-0.722	9.819	6
Class3	-4.57	-1.969	***1.038***	-1.082	1.847	0.519	-1.216	0.228	6.023	4
Class4	***-3.495***	-1.278	1.357	-0.598	2.631	1.174	-0.232	1.067	1.118	1
Class5	***-4.556***	-2.092	0.86	-1.303	1.652	0.387	-1.655	-0.147	7.567	5
Class6	-4.215	-1.915	***0.982***	-1.022	2.044	0.754	-1.178	0.277	4.978	3

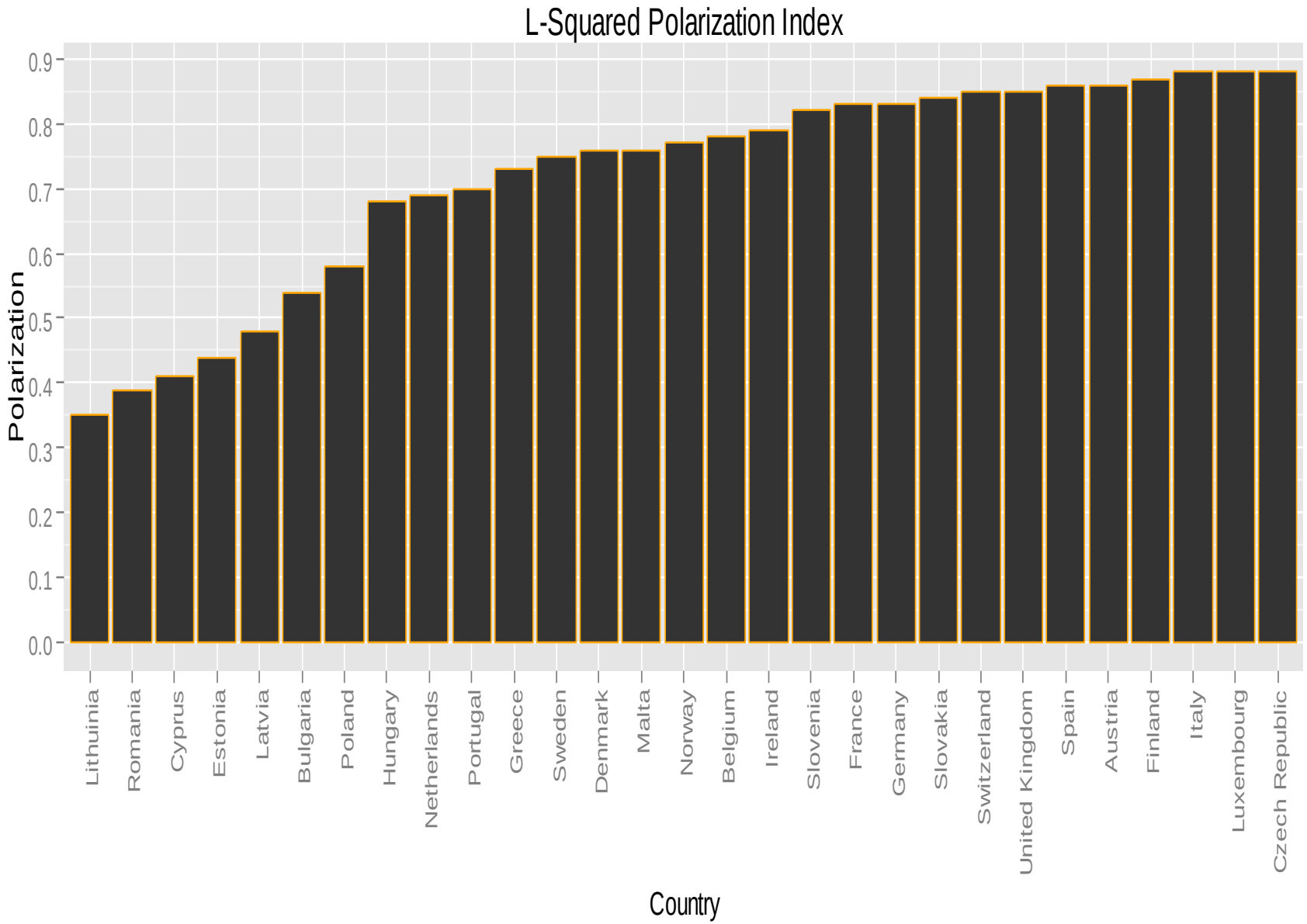
Thresholds and Means Estimates for the Seven-Class Unconstrained Model

	Happy1	Happy2	Happy3	pmat1	pmat2	Trust	petition1	petition2	Homosex	Order
Class1	***-4.164***	-1.46	***1.490***	-0.711	***2.696***	1.028	-0.53	0.803	2.003	2
Class2	***-3.844***	-2.698	1.335	-0.841	2.349	0.854	-0.808	0.722	3.466	3
Class3	***-4.555***	-2.083	0.865	-1.294	1.661	0.394	-1.637	-0.133	7.553	5
Class4	***4.476***	***-2.424***	0.573	-1.759	1.249	-0.26	-2.401	-0.789	9.996	7
Class5	***-4.566***	***-2.454***	0.752	-1.482	1.501	0.044	-2.003	-0.428	8.997	6
Class6	-3.433	-1.256	***1.341***	-0.583	***2.623***	1.193	-0.194	1.104	1.004	1
Class7	-4.301	-1.931	1	-1.038	1.993	0.69	-1.188	0.265	5.267	4


Van der Eijk's Polarization Index
Polarization
0.0
0.1
0.2
0.3
0.4
0.5
Lithuinia
Romania
Cyprus
Estonia
Latvia
Bulgaria
Poland
Netherlands
Hungary
Portugal
Greece
Norway
Denmark
Ireland
Malta
Sweden
Slovenia
Slovakia
Switzerland
Belgium
Finland
Italy
Luxembourg
Czech Republic
France
United Kingdom
Germany
Spain
Austria
Country


Berry/Mielke's Polarization Index
Polarization
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Lithuinia
Romania
Cyprus
Estonia
Latvia
Bulgaria
Poland
Hungary
Ireland
Malta
Portugal
Slovenia
Greece
Slovakia
Sweden
Czech Republic
Italy
United Kingdom
Austria
Luxembourg
Finland
France
Norway
Spain
Germany
Switzerland
Denmark
Belgium
Netherlands
Country


Leik's Polarization Index
Polarization
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Lithuinia
Romania
Cyprus
Estonia
Latvia
Bulgaria
Poland
Hungary
Sweden
Netherlands
Portugal
Belgium
Greece
Denmark
Malta
France
Norway
Ireland
Germany
United Kingdom
Slovenia
Austria
Slovakia
Spain
Switzerland
Italy
Finland
Luxembourg
Czech Republic
Country


L-Squared Polarization Index
Polarization
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Lithuinia
Romania
Cyprus
Estonia
Latvia
Bulgaria
Poland
Hungary
Netherlands
Portugal
Greece
Sweden
Denmark
Malta
Norway
Belgium
Ireland
Slovenia
France
Germany
Slovakia
Switzerland
United Kingdom
Spain
Austria
Finland
Italy
Luxembourg
Czech Republic
Country

Pairwise Correlations between Polarization Measures

	RQ Index	Berry-Mielke	Lsquared	Polarization	Leik
RQ Index	1	0.35	0.52	0.28	0.55
Berry_Mielke	0.35	1	0.82	0.83	0.79
Lsquared	0.52	0.82	1	0.94	0.99
Polarization	0.28	0.83	0.94	1	0.91
Leik	0.55	0.79	0.99	0.91	1

Further development

- Bayesian LCA
- Testing for local homogeneity in IRT frameworks instead of LCA measurement invariance
- Adding covariates
- Any good advice is highly welcomed!!